Parasitol Res (2002) 88: 583–586
DOI 10.1007/s00436-001-0585-0

SHORT COMMUNICATION

Fadi F. Hamdan · Aisha Mousa · Paula Ribeiro

Codon optimization improves heterologous expression of a *Schistosoma mansoni* cDNA in HEK293 cells

Received: 5 October 2001 / Accepted: 10 December 2001 / Published online: 16 February 2002

Abstract Differences in codon usage can seriously hamper the expression of cloned cDNAs in heterologous systems. In this study, we show that the expression of a cloned *Schistosoma mansoni* cDNA in cultured HEK293 cells was dramatically increased by rewriting a portion of the cDNA according to human preferred codon usage, suggesting that codon optimization is a valuable strategy for improving the heterologous expression of helminth sequences. We further describe a simple modification of a recursive PCR-based method, which allows the rewriting of long stretches of DNA sequence in a single PCR reaction. This method can be used to optimize the codon usage of virtually any DNA from helminths and other parasites.

The recent growth of helminth genomic and EST (expressed sequence tag) sequencing data has increased the need for suitable heterologous expression systems in which to test the function of cloned DNA sequences. Unfortunately, helminth DNAs often express poorly in heterologous environments such as *Escherichia coli* or mammalian cells. This is due, in part, to differences in codon usage, which can significantly decrease translation efficiency and thereby lower protein production in a heterologous system (Hernan et al. 1992; Kane 1995). In our laboratory, we have experienced difficulties with heterologous expression of a recently cloned *Schistosoma mansoni* cDNA (GenBank Accession No. AF031196). The cDNA, which encodes a novel *S. mansoni* G protein-coupled receptor (SmGPCR) (Hamdan et al. 2002), was found to express at very low levels in transiently transfected HEK293 cells. Attempts to improve expression by varying transfection conditions, the expression vector or the expressing cell environment were all unsuccessful. A close examination of the DNA sequence of SmGPCR revealed that a significant number of amino acid residues were encoded by codons that are rarely represented in mammalian cells (Nakamura et al. 1999), in particular, T/CTA (Leu), ATA (Ile) and CGT/A (Arg). Moreover, the analysis showed that a majority (≈70%) of these distinctive SmGPCR codons occurred in the first half of the cDNA sequence, frequently in clusters of two to three consecutive codons, which can further contribute to low protein production in the heterologous environment (Kane 1995). To address this problem, we devised a strategy designed to rewrite the *S. mansoni* cDNA according to mammalian (human) preferred codon usage. Codon optimization has been used successfully to improve heterologous expression of cDNAs from other species (Hernan et al. 1992; Yang et al. 1996; Vervoort et al. 2000) and a number of different experimental protocols have been described, which range from the modification of a few codons (Vervoort et al. 2000) to the extensive rewriting of up to 1,000 bp of DNA (Hale and Thompson 1998). DNA rewriting is typically done by recursive (overlap) PCR (Prodromou and Pearl 1992), a technique in which the desired sequence is produced in vitro from the extension of multiple overlapping oligonucleotide primers. Existing protocols for rewriting long DNA molecules (>600 bp), which are more difficult to synthesize in vitro, require multistep procedures of up to three consecutive rounds of recursive PCR (Hale and Thompson 1998; Te'o et al. 2000). In the present study, we describe a simpler method, which was optimized to rewrite the first 855 bp of the *S. mansoni*

F.F. Hamdan · A. Mousa · P. Ribeiro (✉)
Institute of Parasitology, McGill University,
21, 111 Lakeshore Road, Ste. Anne de Bellevue,
Quebec H9X 3V9, Canada
E-mail: paula_ribeiro@maclan.mcgill.ca
Tel.: +1-514-3987607
Fax: +1-514-3987857

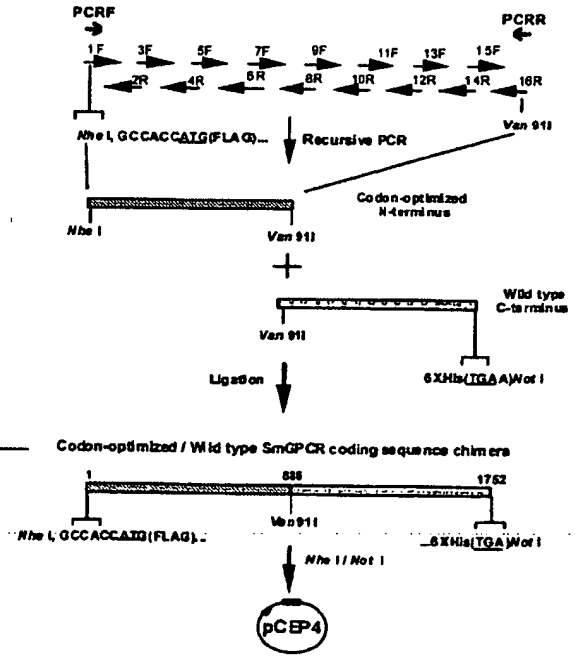
Current address: F.F. Hamdan
Laboratory of Bioorganic Chemistry, Bldg. 8,
NIDDK, National Institutes of Health,
Bethesda, MD 20892, USA

cDNA in a single PCR reaction. The results presented here show that rewriting the first half of the cDNA was sufficient to dramatically increase the expression of full-length SmGPCR protein in cultured HEK293.

The first half of the SmGPCR coding sequence (nucleotides 268–1,122; GenBank Accession No. AF031196) was synthesized by a one-step recursive PCR that contained a total of 16 overlapping mutagenic primers and two external sense (PCRF) and antisense (PCRR) primers. A schematic representation of the recursive PCR strategy and a list of relevant primer sequences are shown in Fig. 1. The mutagenic primers were designed so as to rewrite the *S. mansoni* cDNA using preferred human codons, as defined in the international DNA sequence data base (Nakamura et al. 1999), without changing the amino acid sequence encoded by SmGPCR. The sequence of each mutagenic oligonucleotide was carefully checked to minimize the risk of hairpin formation, which is common in the longer primers, and to avoid the introduction of cryptic splice sites (GTA/TCAG) that may produce truncated products. Adjacent pairs of mutagenic primers overlapped by 19–21 bases so as to allow extension of a continuous DNA product in the recursive PCR reaction. The two sense and antisense external PCR primers (PCRF, PCRR, 25–34 mers) were directed towards the beginning and the end of the codon-optimized SmGPCR sequence, respectively, and were used to amplify the final modified product (Fig. 1). The forward external primer and the first nested mutagenic primer (1F) were designed to introduce a *Nhe*I site followed by a Kozak sequence (GCCACC) (Kozak 1999) immediately preceding the start ATG, and a 24 nucleotide sequence encoding a FLAG epitope (DYKDDDDK) (Hopp et al. 1988). The reverse primer was designed to introduce a flanking *Van*911 site for subsequent ligation of the PCR product into an expression construct (Fig. 1). For the recursive PCR, the 16 mutagenic primers (each at 0.02 μM) and the two external primers (each at 0.8 μM) were combined in a single reaction containing 2.5 units of a proofreading DNA polymerase (PWO), 2 mM $MgSO_4$,

Fig. 1 Construction of a codon-optimized SmGPCR expression plasmid. A nucleotide sequence (positions 268–1,122 of SmGPCR cDNA; GenBank Accession No. AF031196), corresponding to approximately the first half of the SmGPCR coding sequence (Hamdan et al. 2001), was synthesized by recursive PCR, using a total of 16 overlapping mutagenic oligonucleotides and two shorter external PCR primers (PCRF and PCRR). The oligonucleotides were designed to rewrite the cDNA according to human preferred codon usage (Nakamura et al. 1999), as described in the text. Complete primer sequences are shown (*inset*). Forward (sense) and reverse (antisense) oligonucleotides are indicated by *F* and *R*, respectively. The positions of start and stop codons are *underlined*. Restriction sites (*Nhe*I, *Van*911, *Not*I) are shown in *italics*. The Kozak sequence (GCCACC), and the FLAG and oligohexahistidine (6×His) epitopes are shown in *bold*. *CF* and *CR* are the primers used to amplify the wild-type second half of the SmGPCR sequence (nucleotides 1,123–1,950 of SmGPCR cDNA). The two codon-optimized (*hatched box*) and wild-type segments were ligated at a *Van*911 site. The resulting chimeric cDNA was then ligated into a pCEP4 vector at the *Nhe*I/*Not*I sites

1F(80)-5′ ATGGACTACAAGGACGACGACGAC AAGCAGTACATCAACAAGACCAGCCTGAACAGCAGCGT CATCCCCGATAGCCTGAT-3′
2R(75)-5′ CAGGAACAGGGAGATGGTCCACTTGATGATAGGGTTGCTCAGGATCCAGCTCTTGATCAGGC TATGGGGATGAC-3′
3F(80)-5′ TGGACCATCTCCCTGTTCCTGATCATCGCCACTGGCACCACATTCTTCGGCAACCTGCTGAT TATCCTGGCCTTCATCAC-3′
4R(82)-5′ GCAGATCAGCCACGGCGAGGGACACGATGTATTGGTCGGTGATTCTCCTCAGTCTGCTGTTG GTGATGAAGGCCAGGATAAT-3′
5F(85)-5′ CTCGCCGTGGCTGATCTGCTGGTGAGCGTGCTGGTGCTGCCTCTGGCTATCGTGAGACAGAA CCTGGGATATTGGCCATTCGAGA-3′
6R(72)-5′ GGCCATGCACAGCACGATGTTAGCGCTCAGCCAGAATTGACACAGCCTATCGCTCTCGAATG GCCAATATCC-3′
7F(53)-5′ CATCGTGCTGTGCATGGCCTCAATTCTGAACCTGTGTTGCATTAGCCTGGACA-3′
8R(78)-5′ GGCGGTGAACCTGGTCCTCTTTGTAAAATATTTCATGGGTCTAGAGATGGCGATGTATCTGT CCAGGCTAATGCAACA-3′
9F(78)-5′ GAGGACCAGGTTCACCGCCAGCACCATGATCGCCGTCGCCTGGATCCTGCCTCTGATCACCA TGCTGCTCCCCTTCGT-3′
10R(84)-5′ CCTATAAGCCTTGTTGTAGGTGATGTGGCAAGATCCCAGGCCCAGTGTGTGTTGATTGCCG CCCACGAAGGGGAGCAGCATGGT-3′
11F(83)-5′ CCTACAACAAGGCTTATAGGATCTACAGCTCTATCGTCGGATTCTTCGGCCCCTTCCTGCT GATCGCCTACATCTACCTGAGA-3′
12R(84)-5′ GGCTGCTCAGCTTGATGTTGGTGATCTGCAGGACCTTCAGTCTGTGCTTGATGATCCAGAA CACTCTCAGGTAGATGTAGGCGA-3′
13F(71)-CCAACATCAAGCTGAGCAGCCTGAAGAAGCCCAAGTCTCACATCAAGGCCACCAGGAAACCCG CTCCCATC-3′
14R(56)-5′ CCTTGATGTTTTCCCACACCTGCTGGAGGTTGATGATGATGGGAGCGGGTTTCCTG-3′
15F(56)-5′ GTGTGGGAAAACATCAAGGGCAAAATCGGCAAAGTGAACATCTTGAGGAACCAGAG-3′
16R(58)-5′ GTGACCACTATATGGACAGGTGTTCTTAGACTTGCTGCTCTGGTTCCTCAAGATGTTC-3′
PCRF(34)-5′ ATAGCTAGCCACCATGGACTACAAGGACGACGAC-3′
PCRR(25)-5′ GTGACCACTATATGGACAGGTGTTC-3′
CF(37)-5′ AACACCTGT CCATATAGTGGCCATTTTTTCCATTCTG-3′
CR(58)-5′ AAT GCGGCCGCTTCAATGGTGGTGATGGTGGTG CTTATGGTTGTAGCTCTTAATATTC-3′



200 μM of each dNTP and a standard PCR buffer. The cycling protocol was exactly as recommended for the proofreading PWO DNA polymerase (Roche Molecular Biochemicals) at an annealing temperature of 60°C. We found that the success of the recursive PCR reaction was dependent mainly on the ratio of the external PCR primers (PCRF, PCRR) relative to the concentration of internal mutagenic oligonucleotides. Four concentrations of mutagenic primers were tested, 0.005, 0.01, 0.02 and 0.04 μM. A ratio of 0.8 μM of each external primer to 0.02 μM of each mutagenic primer gave the highest specificity and yield.

The remaining half of the SmGPCR coding sequence (1,123–1,950; GenBank Accession No. AF031196) was amplified in a standard PCR, using a sense primer (CF) that introduced a *Van*91I site and overlapped with the 3′-end of the above codon-optimized sequence (Fig. 1, inset). The antisense primer (CR) incorporated a six histidine epitope, followed by a stop codon and a single adenosine immediately after the stop codon (TGAA). The adenosine was added in order to increase translational termination efficiency in the mammalian cell environment (McCaughan et al. 1995). The reverse primer also introduced a *Not*I restriction site for subsequent ligation of the product into an expression vector. To obtain a complete SmGPCR, the codon-optimized and wild-type PCR products were gel purified and digested with *Nhe*I/*Van*91I and *Van*91I/*Not*I, respectively. Subsequently, the two fragments were fused together at the *Van*91I site and then ligated into *Nhe*I/*Not*I sites of digested pCEP4 mammalian expression vector (Invitrogen) (Fig. 1).

pCEP4 expression constructs carrying partially codon-optimized SmGPCR were transiently transfected into HEK293(EBNA1) cells (Invitrogen), using FuGENE6 (Roche Molecular Biochemicals), according to the manufacturer's recommendations. The level of expression of SmGPCR in the transfected HEK293 (EBNA1) cells was monitored by Western blotting and in situ immunofluorescence with antibodies that targeted the N-terminal FLAG epitope and the C-terminal oligohistidine epitope. Among the three codon-optimized clones tested initially, all were found to react with both antibodies, and were thus presumed to be full-length (not shown). One of these clones was subsequently confirmed by DNA sequencing of the entire insert and was selected for further analysis. Figure 2 shows a typical in situ immunofluorescence assay of cells transiently expressing SmGPCR, using an anti-FLAG (M2) antibody (Sigma) followed by a secondary antibody conjugated to fluorescein. The results revealed significant fluorescence in cells transfected with the codon-optimized form of SmGPCR (Fig. 2A, C). Fluorescence was localized mainly in the cell periphery (Fig. 2C), consistent with the *S. mansoni* receptor being expressed at high levels and correctly targeted to the cell membrane. In contrast, a parallel experiment using a wild-type SmGPCR pCEP4 construct, which also carried an N-terminal FLAG epitope, but was not codon-optimized, produced very low fluorescence in transiently transfected HEK293(EBNA1) cells (Fig. 2B). Similar results were obtained from a comparative Western blot analysis of the same codon-optimized and wild-type receptor species (Fig. 2D). Aliquots of a 28,000 *g* crude membrane fraction containing identical amounts of total protein (20 μg) were resolved by sodium dodecyl sulfate

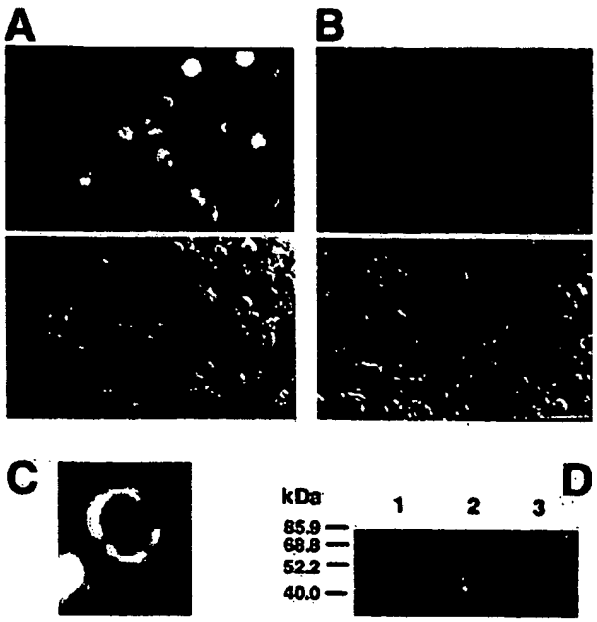


Fig. 2A–D Effect of codon-optimization on SmGPCR expression in HEK293(EBNA1) cells. In situ immunofluorescence analyses: Cells were seeded onto coverslips in six-well plates (2–3×10^5 cells/well), grown to approximately 80% confluency and then transfected with 1 μg of a pCEP4 expression plasmid containing either codon-optimized SmGPCR (A, C) or wild-type SmGPCR (B), each fused at the N-terminal end to a FLAG epitope (Hopp et al. 1988). About 48 h post-transfection, the cells were fixed on ice with 4% paraformaldehyde (Sigma) and then incubated with a monoclonal antibody directed against the FLAG epitope [anti-FLAG(M2), Sigma; 5 μg/ml] followed by a secondary fluorescein isothiocyanate-conjugated antibody (goat anti-mouse IgG; Sigma; 1:2000 dilution). Cells were examined by fluorescence microscopy and conventional visible light, using a Nikon Optiphot-2 fluorescence microscope equipped with Nomarski optics. Transfection efficiency, measured in parallel experiments, was approximately 40% for both populations of cells. C A closer view of a transfected cell shows expression of the codon-optimized form of SmGPCR mainly in the cell periphery. D Western blot analysis: approximately 1–1.5×10^6 HEK293(EBNA1) cells were transfected with 3 μg each of codon-optimized SmGPCR-FLAG (*lane 1*), wild-type SmGPCR-FLAG (*lane 2*), or pCEP4 plasmid alone without insert (*lane 3*). Cells were lysed 48 h post-transfection and a crude 28,000 *g* membrane fraction was prepared from each population of cells. Aliquots of the three membrane preparations containing the same amount of total protein (20 μg) were resolved on 12% sodium dodecyl sulfate polyacrylamide gels and subsequently electroblotted onto nitrocellulose filters (Novex). Filters were incubated with anti-FLAG(M2) antibody (1:1000 dilution) followed by a horseradish peroxidase-conjugated secondary antibody (1:2000 dilution), according to standard procedures. Western positive bands were detected on film using Lumilight (Roche). The positions of relevant protein standards are indicated

polyacrylamide gel electrophoresis and immunoblotted with the anti-FLAG (M2) antibody, according to standard protocols. A strong immunoreactive band corresponding to the SmGPCR monomer (approximately 65 kDa) was observed in samples transfected with codon-optimized receptor (lane 1) whereas the wild-type produced a much weaker, barely detectable band (lane 2). No signal was detected in the 28,000 *g* supernatant of either transfection (not shown) or in membranes of control cells transfected with pCEP4 plasmid only (lane 3). Taken together, the results clearly demonstrate that the codon-rewriting strategy was successful in increasing SmGPCR expression in the mammalian cell.

In conclusion, this study describes a relatively simple, PCR-based codon-optimization procedure that can be used to improve the translation efficiency of any cDNA in a heterologous environment. It is noteworthy that this strategy was successful without having to rewrite the entire SmGPCR cDNA sequence, which would have complicated the recursive PCR further. Rewriting only the first half of SmGPCR, which contained the highest proportion of rare codon clusters, was sufficient to increase protein expression, presumably because the lower frequency of rare codons allowed for higher translation efficiency in the HEK293 cells. The same strategy could be applied to rewrite any other region of a cDNA of interest, at least up to 855 bp and, possibly, even longer sequences. This method is potentially useful for heterologous expression of cDNAs from helminths as well as other parasites, including protozoans, which have distinctive codon preferences.

Acknowledgements We thank Joseph Nabhan for excellent assistance and Dr. R. Sinibaldi (OPERON) and Dr. Y. Durocher (Biotechnology Research Institute) for useful discussions on codon optimization. This work was supported by a grant from the Canadian Institutes of Health Research (CIHR) to P.R.

References

Hale RS, Thompson G (1998) Codon optimization of the gene encoding a domain from human type 1 neurofibromin protein results in a threefold improvement in expression level in *Escherichia coli*. Protein Expr Purif 12:185–188

Hamdan FF, Abramovitz M, Mousa A, Xie J, Durocher Y, Ribeiro P (2002) A novel *Schistosoma mansoni* G protein-coupled receptor is responsive to histamine. Mol Biochem Parasitol 119:75–86

Hernan RA, Hui HL, Andracki ME, Noble RW, Sligar SG, Walder JA, Walder RY (1992) Human hemoglobin expression in *Escherichia coli*: importance of optimal codon usage. Biochemistry 31:8619–8628

Hopp TP, Prickett KS, Price VL, Libby RT, March CJ, Cerretti DP, Urdal DL, Conlon PJ (1988) A short polypeptide marker sequence useful for recombinant protein identification and purification. Biotechnology 6:1204–1210

Kane JF (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. Curr Opin Biotechnol 6:494–500

Kozak M (1999) Initiation of translation in prokaryotes and eukaryotes. Gene 234:187–208

McCaughan KK, Brown CM, Dalphin ME, Berry MJ, Tate WP (1995) Translational termination efficiency in mammals is influenced by the base following the stop codon. Proc Natl Acad Sci U S A 92:5431–5435

Nakamura Y, Gojobori T, Ikemura T (1999) Codon usage tabulated from the international DNA sequence databases; its status 1999. Nucleic Acids Res 27:292

Prodromou C, Pearl LH (1992) Recursive PCR: a novel technique for total gene synthesis. Protein Eng 5:827–829

Te'o VS, Cziferszky AE, Bergquist PL, Nevalainen KM (2000) Codon optimization of xylanase gene xynB from the thermophilic bacterium *Dictyoglomus thermophilum* for expression in the filamentous fungus *Trichoderma reesei*. FEMS Microbiol Lett 190:13–19

Vervoort EB, van Ravestein A, van Peij NN, Heikoop JC, van Haastert PJ, Verheijden GF, Linskens MH (2000) Optimizing heterologous expression in *Dictyostelium*: importance of 5′ codon adaptation. Nucleic Acids Res 28:2069–2074

Yang TT, Cheng L, Kain SR (1996) Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein. Nucleic Acids Res 24:4592–4593